

Wild Cross-validation for Density Estimation

Olga Y. Savchuk, Jeffrey D. Hart, Simon J. Sheather
Department of Statistics, Texas A&M University

Abstract: A new method of selecting the bandwidth of a kernel density estimator is proposed. The method, termed wild cross-validation, uses least squares cross-validation (LSCV) to select the bandwidth of a so-called wild kernel, and then rescales this bandwidth to be appropriate for use in a Gaussian kernel estimator. The wild kernels are linear combinations of two Gaussian kernels, and are wild since they need not be unimodal or positive. We develop theory showing that the relative error of wild CV bandwidths can converge to 0 at a rate of $n^{71/4}$, which is substantially better than the $n^{71/10}$ rate of LSCV. The wild CV method uniformly outperforms LSCV in a simulation study and in two real data examples.

Keywords: Cross-validation; Bandwidth selection; Kernel density estimation.